

# Prediction of US election with linear model\*

Colin Sihan Yang      Lexun Yu      Siddharth Gowda

October 21, 2024

This paper forecast the winner of the upcoming US presidential election using “poll-of-polls” by building a linear model.

## 1 Introduction

## 2 Data

### 2.1 Overview

### 2.2 Model

$$\begin{aligned} \text{candidate\_percentage} = & -0.218 \times \text{transparency\_score} + 0.321 \times \text{pollscore} \\ & + 0 \times \text{sample\_size} + \text{candidate\_intercept} \times \text{candidate\_name\_vector} \\ & + 51.371 \end{aligned} \tag{1}$$

From Table 2, candiate name is the biggest factor in terms of support. The name was added to the model to give a baseline not to show any correlation. After that, poll score and transparency score are the biggest factors. These scores are two baselines of how reliable and trustable the pollsters are. The sample size had a coefficent of zero, meaning it had no affect on the model’s output.

The model also has a very high  $R^2$  value in 0.962 Most of this can be attributed to using the candidate name (anwser) as a variable.

Table 1: Coefficients for Multilinear Regression Model

	(1)
(Intercept)	51.371 (0.787)
transparency_score	−0.218 (0.084)
pollscore	0.321 (0.308)
sample_size	0.000 (0.000)
answerKennedy	−43.400 (0.424)
answerStein	−47.155 (0.306)
answerTrump	−1.900 (0.233)
days_taken_from_election	−0.018 (0.004)
Num.Obs.	1386
R2	0.962
R2 Adj.	0.962
AIC	7649.2
BIC	7696.3
Log.Lik.	−3815.596
RMSE	3.80

Table 2: Coefficients for Multilinear Regression Model

[htbp]	
Variable	Coefficient
Intercept	51.371
Transparency Score	-0.218
Poll Score	0.321
Sample Size	0.000
Candidate Name	Varies
Days from Election	Not shown

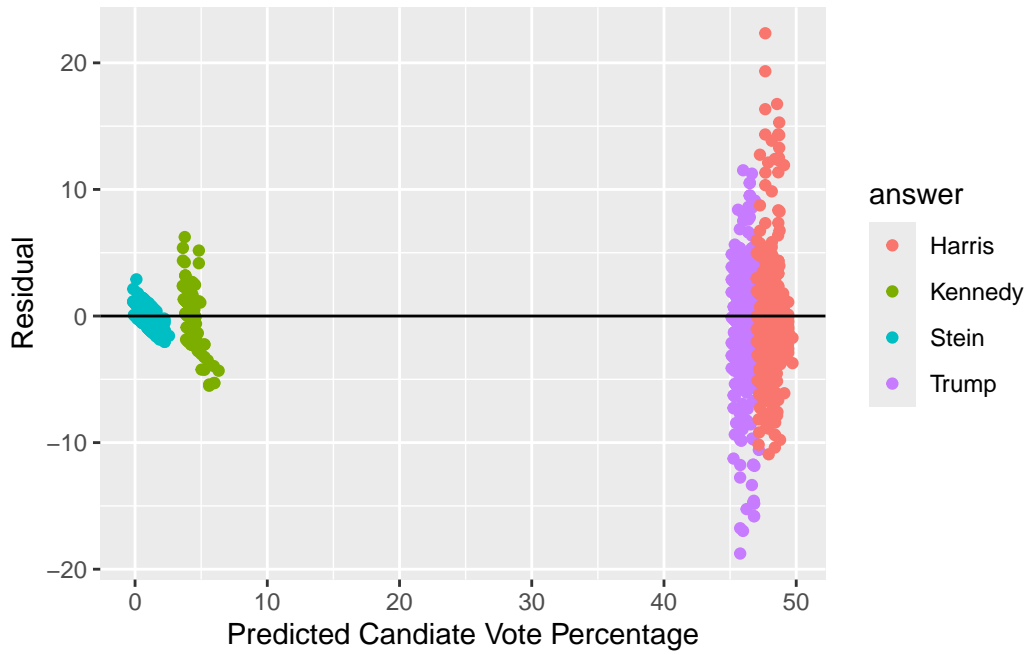


Figure 1: Multilinear Regression Model Residuals

## 2.3 Residuals

Based on Figure 1, there appears to be no patterns in the residuals. This means that a linear model like this could be appropriate. However, there does seem to be a big difference in terms of residual variability based on candidate name.

---

\*Code and data are available at: <https://github.com/yulexun/uselection>.

### 3 References